

Low Exchangeability of Selenocysteine, the 21st Amino Acid, in Vertebrate Proteins

Sergi Castellano,^{*1} Aida M. Andrés,^{*1} Elena Bosch,[†] Mònica Bayes,^{‡§} Roderic Guigó,[‡] and Andrew G. Clark^{*}

^{*}Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY; [†]Departament de Ciències Experimentals i de la Salut, Institut de Biologia Evolutiva (UPF-CSIC), Universitat Pompeu Fabra, Barcelona, Spain; [‡]Centre de Regulació Genòmica, Barcelona, Spain; and [§]Centro Nacional de Genotipado (CeGen), Barcelona, Spain

Selenocysteine (Sec), the 21st amino acid, is incorporated into proteins through the recoding of a termination codon, an inefficient translational process mediated by a complex molecular machinery. Sec is a rare amino acid in extant proteins, chemically similar to cysteine (Cys), found in homologous position to Cys of nonselenoprotein families. Selenoproteins account for the dependence of vertebrates on environmental selenium (Se) and have an important role in several Se-deficiency diseases. Selenoproteins are poorly characterized enzymes and reports on the functional exchangeability of Sec with Cys are limited and controversial. Whether the unique role of Sec in some selenoenzymes illustrates the broader contribution of Se to protein function is unknown (Gromer S, Johansson L, Bauer H, Arscott LD, Rauch S, Ballou DP, Williams CH Jr, Schirmer RH, Arnér ES. 2003. Active sites of thioredoxin reductases: why selenoproteins? *Proc Natl Acad Sci USA*. 100:12618–12623). Here, we address this question from an evolutionary perspective by the simultaneous identification of the patterns of divergence in almost half a billion years of vertebrate evolution and diversity within the human lineage for the full complement of enzymatic Sec residues in these proteomes. We complete this analysis with data for the homologous Cys residues in the same genomes. Our results indicate concerted purifying selection across Sec and Cys sites in all selenoproteomes, consistent with a unique role of Sec in protein function, low exchangeability, and an unknown degree of functional divergence with Cys homologs. The distinct biochemical properties of Sec, rather than the geographical distribution of Se, global O₂ levels or Sec metabolic cost, appear to play a major role in driving adaptive changes in vertebrate selenoproteomes. A better understanding of the selenoproteomes and neutral evolutionary patterns in other taxa will be necessary to fully assess the generality of this conclusion.

Introduction

Patterns of amino acid changes in proteins are often interpreted as a measure of the exchangeability between amino acid pairs. Indeed, the propensity for evolutionary change from one amino acid to another has been thoroughly studied. For example, the AAindex database (Kawashima et al. 2008) contains dozens of exchangeability measures between amino acids, from purely physicochemical to purely comparative. A functional interpretation of these measures is common in evolutionary studies. This research, however, applies only to the standard 20 amino acids in proteins. Little is known about the exchangeability of selenocysteine (Sec), the 21st amino acid, in nature.

Sec is a cysteine (Cys) analogue with a selenium (Se)-containing selenol group in place of the sulfur-containing thiol group in Cys (Stadtman 1996). Sec and Cys residues occupy homologous positions, presumably serving an oxidoreductase role, in several proteins not fully functionally characterized. Although the translation of Cys codons (UGC/UGT) in these proteins is typical, a complex translational machinery is necessary to incorporate Sec into an in-frame termination codon (UGA) in selenoproteins (Driscoll and Copeland 2003). In eukaryotes, the Sec insertion sequence (SECIS), an RNA stem-loop located in the 3' untranslated region of selenoprotein genes, recruits several transacting factors to recode UGA from termination to Sec

insertion (Driscoll and Copeland 2003). The Sec residue encoded by the recoded UGA is inserted into the growing peptide, and translation of the protein continues until the proper termination codon.

The majority of eukaryotic and prokaryotic selenoproteins have now been found in Cys form raising questions about the functional exchangeability of Sec with Cys in protein function, a long-standing issue in Se biology (Johansson et al. 2005). Among those clades with Se-dependent proteins, the growing knowledge of the number, distribution, and function of vertebrate selenoproteins and their Cys-containing homologs (Gromer et al. 2003; Kryukov et al. 2003; Castellano et al. 2004, 2005; Kim and Gladyshev 2005; Shchedrina et al. 2007), together with the evolutionary depth and quality of vertebrate genomes, provides a first opportunity to gain insight into this question.

The extent of exchangeability between Sec and Cys residues reflects the contribution of Sec to protein function. The long-term exchangeability between these amino acids, however, cannot be fully ascertained from functional studies on extant selenoproteins, as functional differences in present-day sequences are not a measure of fitness in natural populations (Gould and Lewontin 1979; Eyre-Walker and Keightley 2007; Nielsen et al. 2007). Therefore, the best approach to evaluate the functional exchangeability between the two amino acids is to infer the strength and mode of natural selection acting on the reciprocal exchange of residues (Williamson et al. 2005), as amino acid exchangeability is a measure of the neutrality of the substitution of two residues by one another in a protein.

Neutral patterns of selenoproteome divergence and diversity would indicate no fitness advantage or disadvantage of Sec over Cys (e.g., no distinct contribution of Sec to protein activity). Under neutrality, the expected patterns of variation in Sec and Cys sites can be inferred from population genetics

¹ Author contribution: S.C. and A.M.A. contributed equally to this work.

Key words: selenium, selenocysteine, cysteine, selenoproteins, vertebrates, exchangeability.

E-mail: castellanos@janelia.hhmi.org.

Mol. Biol. Evol. 26(9):2031–2040. 2009

doi:10.1093/molbev/msp109

Advance Access publication June 1, 2009

theory or simulations of the evolutionary process (Castellano 2009). This constitutes the null (undisturbed by selective forces) model of Sec usage in protein evolution. Departures from neutrality are a signature of natural selection and, under some simplifying assumptions, can be interpreted as 1) selection against deleterious Sec/Cys mutations (purifying selection), which is consistent with low Sec/Cys exchangeability and denotes functional differences between Sec and Cys residues; purifying selection would result in a deficit of variation within populations and differentiation among populations, or 2) selection favoring advantageous Sec/Cys mutations (positive selection), which can be interpreted as evidence for adaptive evolution and, in the case of heterogeneous selective pressures unrelated to protein function, high Sec/Cys functional exchangeability; positive selection in some populations would result in an overall excess of variation.

Environmental, metabolic and biochemical selective pressures have been suggested to shape Sec use in proteins. Among those suggested are 1) the wide differences in dietary Se status among populations due to the worldwide variability of Se content in soils and waters (Shamberger 1981; Levander 1987; Valentine 1997), which may lead to disease due to excess or deficit of Se (Levander 1987); 2) the different Sec sensitivities to oxidation among selenoproteins and selenoproteomes due to variable O₂ levels over geologic time (Leinfelder et al. 1988; Jukes 1990; Berner 2006; Berner et al. 2007); 3) the higher anabolic cost and lower translational efficiency of Sec (Berry et al. 1992; Driscoll and Copeland 2003; Mehta et al. 2004; Xu et al. 2007); and 4) the increased reactivity provided by Se (Berry et al. 1992; Rocher et al. 1992; Maiorino et al. 1995; Zhong and Holmgren 2000), which results in a possibly advantageous high catalytic activity in selenoenzymes (100- to 1000-fold more active than their Cys counterparts). Such higher enzymatic efficiency of Sec over Cys has, however, recently been challenged (Kanzok et al. 2001; Gromer et al. 2003; Kim and Gladyshev 2005) and its interpretation in terms of functional exchangeability over evolutionary time is problematic. The importance of these selective factors in selenoprotein evolution is untested, despite their common explanatory role for Sec/Cys replacements in the Se and selenoprotein literature.

Here, we study the exchangeability between Sec and Cys enzymatic residues through the analysis of the patterns of divergence among vertebrates and diversity within humans of all homologous Sec and Cys sites in these genomes. We identify concerted purifying selection across these sites in all selenoproteomes, consistent with a unique role of Sec in protein activity. The low exchangeability observed between Sec and Cys amino acids reveals a previously unappreciated degree of functional divergence between Sec- and Cys-containing enzymes where the distinct biochemical properties of Sec, and not environmental nor metabolic factors, may drive adaptive changes in selenoproteomes. Our conclusions represent a strong departure from the recent but prevailing view favoring ecological explanations to Sec evolution.

Materials and Methods

The Human Selenoproteome

The human selenoproteome consists of 25 selenoproteins (Kryukov et al. 2003) and 6 paralogous genes with

Cys (supplementary table S1, Supplementary Material online). In addition, four genes with Cys, orthologous to vertebrate selenoproteins, exist (supplementary table S1, Supplementary Material online). All 35 proteins were included in the divergence and diversity analyses. However, only the enzymatic Sec residue in the N-terminal domain of SelP was analyzed. Human sequences were obtained from SelenoDB (Castellano et al. 2008) at <http://www.selenodb.org>. This database provides the correct genomic structure of all human selenoprotein genes, which is essential to our genotyping efforts. Other more general databases only contain the mRNA sequence (e.g., Genbank) or vertebrate selenoprotein gene annotations of variable quality (e.g., Ensembl) and are not appropriate for this study.

Nonhuman Selenoproteomes

For completeness, we included these genes in our divergence analysis (supplementary table S2 and supplementary fig. S1, Supplementary Material online).

Orthology Assignment

Following Nikolaev et al. (2007) phylogeny, one or more representative from all major vertebrate taxa were chosen 1) Supraprimates: Human, Chimpanzee, Macaque, and Mouse; 2) Laurasiatheria: Dog and Hedgehog; 3) Xenarthra: Armadillo; 4) Afrotheria: Elephant; 5) Marsupialia: Opossum; 6) Prototheria (Monotremata): Platypus; 7) Archosauromorpha (crocodiles, dinosaurs, and birds): Chicken; 8) Lepidosauromorpha (snakes and lizards): Lizard; 9) Amphibia (salamanders and frogs): Frog; and 10) Teleostei: Puffer fish and Zebrafish. The phylogeny encompasses 450 ± 36 My of vertebrate evolution (Hedges 2002; fig. 1). See supplementary table S3, Supplementary Material online for species scientific names.

Nonhuman selenoproteins are routinely misannotated in genomic projects as most gene annotation systems consider Sec as a stop codon. The correct gene structures and protein sequences for most selenoproteins in diverged species are not easily available and their annotation will involve extensive manual curation in the future. Therefore, orthologous residues to 25 Sec and 10 homologous Cys amino acids in human were identified as follows: 1) The 35 human proteins, organized into 19 families (supplementary table S1 in Supplementary Material online), were blasted (Gish and States 1993) against the panel of vertebrate genomes. WU Blast 2.0 parameters were $E = 0.001$, $W = 4$, and filter = seg in combination of the substitution matrices BLOSUM50, 62 or 80; 2) all hits were automatically filtered for alignments with symmetrical conservation in regions flanking Sec–Sec or Cys–Sec aligned pairs (at least 5 similar residues in both regions of 10 amino acids each); and 3) the target sequences in each alignment were blasted back against the human families. Orthology was assigned by best reciprocal hit. In each step, alignments were manually inspected and extended beyond the Sec codon if necessary. Thioredoxin-like proteins were searched without the symmetrical conservation filter due to the short sequence region beyond Sec. When

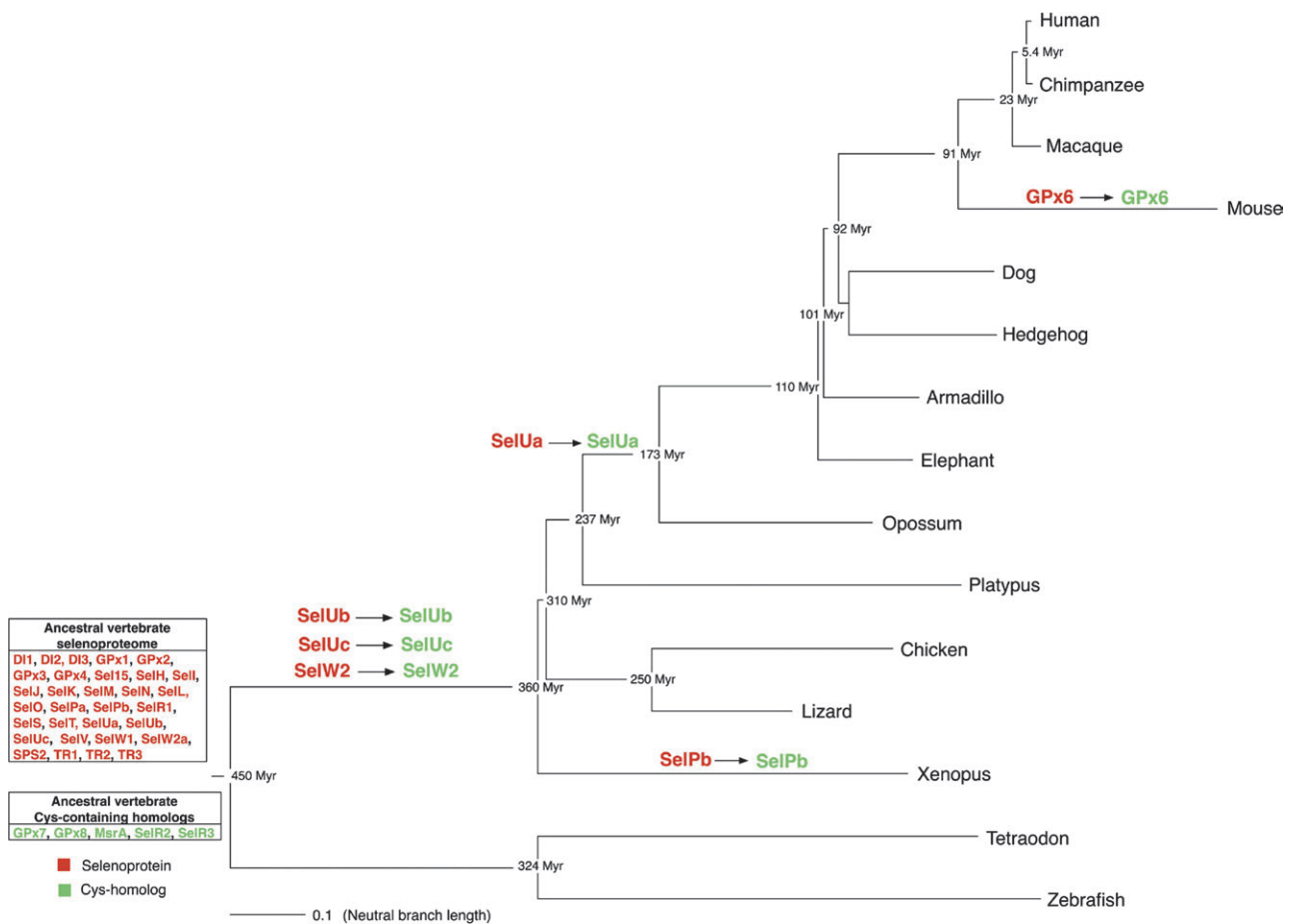


FIG. 1.—Divergent evolution of vertebrate selenoproteomes illustrated with the inferred ancestral vertebrate selenoproteome, ancestral Cys homologs, and Sec/Cys changes between orthologous genes along the phylogeny.

available, shark, lamprey, and sea urchin sequences were used to polarize Sec/Cys states. Gene orthology and paralogy were identified through gene and species tree reconciliation (Zmasek and Eddy 2001). Orthology assignment for nonhuman or nonmammalian selenoproteins was carried out similarly.

Reconstruction of Ancestral Selenoproteomes

The program Mesquite v1.12 (<http://mesquiteproject.org>) was used to assign optimal character states (Sec/Cys) to the proteins in the tree internal nodes using the most-parsimonious reconstruction under the Fitch parsimony criterion (Fitch 1971). The overall similarity of selenoproteomes observed among species and the small number and phylogenetic distribution of identified Sec/Cys changes makes the use of Maximum Parsimony a reasonable choice.

Ancestral Selenoproteome Size and O₂ Levels

Twelve reconstructed ancestral selenoproteome sizes throughout the vertebrate phylogeny (fig. 1) were correlated with the estimated levels of atmospheric O₂ (Berner 2006; Berner et al. 2007; supplementary table S4, Supplementary

Material online). To compute Spearman's rank correlation (no normality in the data assumed) given ties between ranks, we calculated Pearson's rank correlation coefficient with the `cor.test` function of the R statistics package v2.5.1. We tested significance with a one-sided *t*-test using the same program.

Divergence Analysis of Enzymatic Sec/Cys Sites

The neutrality test is based on a simple divergence statistic *D* for binary Sec/Cys data. Let $D_{\text{Sec} \rightarrow \text{aa}}$ be the proportion of ancestral Sec sites that have diverged at least once from the ancestral Sec state in a phylogeny. Similarly, let $D_{\text{Cys} \rightarrow \text{aa}}$ be the proportion of ancestral Cys sites that have diverged at least once from the ancestral Cys state in a phylogeny. Then, over *n* ancestral sites of the same class, *D* is simply computed as:

$$D = \frac{1}{n} \sum_{i=1}^n d_i,$$

where $d_i = 0$ if the *i*th site has not diverged in any species from its ancestral state, and $d_i = 1$ if the *i*th site has diverged in one or more species from its ancestral state. This expression of divergence is highly conservative to deviations

toward purifying selection because multiple changes in a site (expected under neutrality) do not affect (counted as one) the estimate of D .

The neutrality test is carried out comparing the empirically observed D_{Obs} with the distribution of D , which was obtained through neutral simulations of the evolution of ancestral Sec or Cys codons along the phylogeny (Nikolaev et al. 2007; fig. 1). We use a continuous time Markov Chain model of sequence evolution and assume independence between sites, a reasonable assumption because there is only one Sec or Cys codon per gene. We performed 10,000 Markov Chain Monte Carlo simulations with a modified version of the program Seq-Gen v1.3.2 (Rambaut and Grassly 1997), in which strongly deleterious mutations (those resulting in TAA or TAG stop codons) are immediately eliminated from the population and do not contribute to sequence divergence (source code and program available from SC). We used the standard Hasegawa-Kishino-Yano model of nucleotide evolution (Hasegawa et al. 1985) in which the instantaneous rate of evolution is comprised of a transition/transversion ratio (TS/TV) set to 1.8 (Rosenberg et al. 2003) and the equilibrium nucleotide frequencies set to $A = 0.26$, $T = 0.26$, $C = 0.24$, and $G = 0.24$, as estimated from 4-fold degenerate sites in 10 vertebrate species ranging from human to *Takifugu* (Margulies et al. 2005). Branch lengths were set to a proxy of the mean number of neutral mutations per site, the number of synonymous substitutions per site, as estimated from 4-fold degenerate sites (Margulies et al. 2005) for the set of species analyzed in this work (Margulies EH, personal communication).

The exchangeability test is also based on the statistic D . Let $D_{\text{Sec} \rightarrow \text{Cys}}$ be the proportion of ancestral Sec sites that have diverged to Cys at least once from the ancestral Sec state in a phylogeny. Similarly, let $D_{\text{Cys} \rightarrow \text{Sec}}$ be the proportion of ancestral Cys sites that have diverged to Sec at least once from the ancestral Cys state in a phylogeny. The exchangeability test is carried out comparing the observed D_{Obs} with the neutral distribution of D , which is derived from 10,000 simulations where mutations other than between Sec and Cys codons are effectively removed from the population by strong purifying selection (see above for mutational parameters). In this model, sequences evolve neutrally at a fraction of the mutation rate. Only Sec/Cys neutral substitutions contribute to selenoproteomes divergence and, therefore, this test reflects the overall proportion of deleterious Sec/Cys (probably function-altering) substitutions removed by the action of purifying selection.

The distribution of mutation rate heterogeneity at the megabase scale is difficult to estimate, as it depends on chromosomal position, GC content, neighboring bases, efficiency of the repair system, and other factors (Ellegren et al. 2003). We investigated the robustness of our tests to nonuniform mutation rates in selenoprotein genes through simulations of the neutral process with increasing rate heterogeneity. We carried out simulations in which mutation rates among genes obey a gamma distribution with decreasing values of the shape (alpha) parameter so that approximately 10%, 15%, 20%, and 25% of the genes lie in mutation cold spots (see Results and supplementary fig. S2, Supplementary Material online).

Human Samples, Genotyping, and Sequencing

The Human Genome Diversity Panel–Centre d'Etude du Polymorphisme Humain (HGDP–CEPH) Human Genome Diversity Cell Line panel contains 1,037 samples from a wide range of world populations (Cann et al. 2002; supplementary table S5, Supplementary Material online). In order to assess variability at 25 Sec (TGA) and 10 Cys (TGT/TGC) enzymatic codons in the panel, SNPLEX genotyping assays (Applied Biosystems, Foster City, CA) for all sites and KASPar assays (Kbioscience, United Kingdom) for 14 codons (unresolved with the previous method) were designed. For each codon, two possible changes were tested: TGA to TGC/TGC, TGT to TGC/TGA, and TGC to TGT/TGA. SNPLEX and KASPar genotyping assays were performed according to manufacturer conditions and average call rates of 98.34% and 98.4% were obtained, respectively (supplementary table S6, Supplementary Material online). Unclear genotypes (8 codons in a total of 11 individuals) were confirmed through direct sequencing. Regions flanking each codon were amplified in a 10- μ l final volume containing PCRx Amplification Buffer (Invitrogen), 1.5 mM MgSO_4 , 2 \times Enhancer Solution (PCRx Enhancer System, Invitrogen), 0.5 μ M of each primer (supplementary table S7, Supplementary Material online), 200 μ M deoxyribonucleotide triphosphate, 0.5 U Taq DNA polymerase (Roche), and 1 μ l of 10–50 ng of DNA. Polymerase chain reaction (PCR) cycling conditions were as follows: 95 °C for 5 min; 32 cycles of 95 °C for 30 s, 46.0 °C (GPx3), 48.0 °C (GPx8), 48.0 °C (MsrA), 52.0 °C (SelH), 51.4 °C (SelM), 50.7 °C (SelN), 48.0 °C (SelT), 48.7 °C (TR1) for 30 s, and 72 °C for 2 min; with a final elongation step at 72 °C for 7 min. Amplification products were purified with Exo-SAP (GE Healthcare Europe GMBH). Sequencing reactions were performed for each strand, using the corresponding forward or reverse amplification primer with the ABI PRISM Big-Dye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) following the supplier's instructions. Sequencing products were subsequently purified using the Montage SEQ96 clean up kit (Millipore) and analyzed on a 3730XL DNA Analyzer (Applied Biosystems).

Diversity Analysis of Enzymatic Sec/Cys Sites

The survey of variability revealed no variation in Sec and homologous Cys sites in humans. The probability of finding, under neutrality, no variation in 35 sites and 2,074 chromosomes given the average level of diversity in humans was assessed based on coalescent theory: first, analytically (Hein et al. 2005), considering the average θ of the human genome (8.25×10^{-4}) (Venter et al. 2001) and second, through coalescent simulations. Neutral coalescent simulations simulate the expected diversity of a locus under neutrality given its mutation rate, sample size, and the demographic history of the population. The evolution of the 35 sites was simulated with the program ms (Hudson 2002), considering the above θ in the human genome, under demographic equilibrium and a range of demographic scenarios compatible with the history of human populations (Marth et al. 2004; Williamson et al. 2005). Though all demographic models represent simplified versions of the

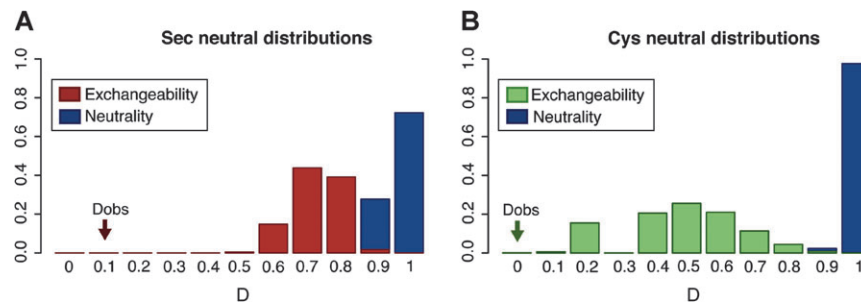


FIG. 2.—Expected distributions under neutrality of the test statistic D . The observed D statistic in vertebrates, indicated as D_{Obs} , has the same value in the neutrality and exchangeability tests (no changes other than between Sec and Cys were found). The distributions used in each test are depicted together in each panel. (A) Neutral distribution of the proportion of divergent Sec sites ($D_{\text{Sec} \rightarrow \text{aa}}$), in blue, and neutral distribution of the proportion of divergent Sec sites ($D_{\text{Sec} \rightarrow \text{Cys}}$) in the Sec/Cys exchangeability test, in red; (B) Neutral distribution of the proportion of divergent Cys sites ($D_{\text{Cys} \rightarrow \text{aa}}$), in blue, and neutral distribution of the proportion of divergent Cys sites ($D_{\text{Cys} \rightarrow \text{Sec}}$) in the Sec/Cys exchangeability test, in green. Due to the small number of simulated Cys sites ($n = 10$), the discrete $D_{\text{Cys} \rightarrow \text{Sec}}$ distribution does not contain values in all bins.

complex demographic history of human populations, fitting demographic parameters in a model of the demographic history of 51 populations is not practical given the small samples in each. Ignoring the population structure present in the data makes the test conservative, because population differentiation increases variability and reduces our power to reject neutrality. Ten thousand simulations were performed for every model, and the probability of the data was assessed by direct comparison of the number of observed polymorphic sites with the simulated distribution. Analytical and simulation results agree, regardless of the demographic model considered.

Results

Reconstruction of Ancestral Selenoproteomes

Our reconstruction of the ancestral vertebrate selenoproteome resulted in 31 selenoproteins (DI1, DI2, DI3, GPx1, GPx2, GPx3, GPx4, Sel15, SelH, SelI, SelJ, SelK, SelM, SelN, SelL, SelO, SelPa, SelPb, SelR1, SelS, SelT, SelUa, SelUb, SelUc, SelV, SelW1, SelW2a, SPS2, TR1, TR2, and TR3) and 5 Cys homologs (GPx7, GPx8, MsrA, SelR2, and SelR3). See figure 1 and supplementary tables S1 and S2, Supplementary Material online. In addition, six non-ancestral selenoproteins (DI4, Fep15, GPx6, SelT2, SelT3, and SelR4) and one Cys homolog (GPx5) exist (supplementary fig. S1, Supplementary Material online). Three non-ancestral selenoprotein genes (SelT3, SelR4, and DI4) occur in a single species and provide no divergence information (supplementary fig. S1, Supplementary Material online). Therefore, the divergence of 35 Sec (SelL has two Sec codons) throughout the phylogeny is considered. Only six Sec/Cys replacements were found in the whole vertebrate tree, all leading to the loss of Sec (fig. 1). Four of these changes to Cys occur deep within the vertebrate phylogeny, and their subsequent divergence can be simulated in the corresponding clades. Together with the 6 Cys sites at the root of the tree, the divergence of a total of 10 Cys sites is, then, simulated.

Ancestral Selenoproteome Size and O_2 Levels

To test whether Sec has an increased sensitivity to rising O_2 levels in vertebrate evolution (Berner 2006; Berner et al. 2007), we computed the correlation of the inferred ancestral

selenoproteome sizes along the internal nodes in the phylogeny with the estimated levels of atmospheric O_2 in the corresponding geological periods (see Materials and Methods). The predicted negative correlation, at the 5% significance level, between selenoproteome size and environmental O_2 levels is not significant ($r = -0.09516$, $P = 0.38430$).

Divergence Analysis of Enzymatic Sec/Cys Sites

Our neutrality test on the pattern of divergence of all enzymatic Sec ($D_{\text{Sec} \rightarrow \text{aa}} = 0.171$, $P < 0.0001$ and homologous Cys ($D_{\text{Cys} \rightarrow \text{aa}} = 0$, $P < 0.0001$) sites over half a billion years of vertebrate evolution is consistent with strong purifying selection (fig. 2A, $D_{\text{Sec} \rightarrow \text{aa}}$ in blue, and B, $D_{\text{Cys} \rightarrow \text{aa}}$ in blue). This experiment, though, informs us of the extent of constraint acting on these sites, not of the functional exchangeability between Sec and Cys residues. To test for this, we introduce an exchangeability test (see Materials and Methods) that takes into account the chemical analogy and evolutionary homology of Sec and Cys residues in the context of pervasive purifying selection for any other amino acid substitution (i.e., other than Sec/Cys replacements). In practice, we derive a null model through simulations of the evolutionary process where strongly deleterious mutations (with strongly negative fitness effects) are immediately eliminated from the population and only Sec/Cys neutral substitutions contribute to selenoproteomes divergence. This test shows a deficit of divergence when compared with the neutral expectation, which is consistent with strong purifying selection and indicates very low exchangeability between Sec ($D_{\text{Sec} \rightarrow \text{Cys}} = 0.171$, $P < 0.0001$) and Cys ($D_{\text{Cys} \rightarrow \text{Sec}} = 0$, $P = 0.0008$) residues in vertebrate proteins (fig. 2A, $D_{\text{Sec} \rightarrow \text{Cys}}$ in red, and B, $D_{\text{Cys} \rightarrow \text{Sec}}$ in green). Although $D_{\text{Sec} \rightarrow \text{aa}}$ and $D_{\text{Sec} \rightarrow \text{Cys}}$ are different evolutionary measures, they have the same value in vertebrates because only changes between Sec and Cys residues were found. Both tests are robust to mutation heterogeneity across a genome (see Discussion and supplementary fig. S2, Supplementary Material online).

Diversity Analysis of Enzymatic Sec/Cys Sites

We further investigated the role of environmental Se in driving Sec/Cys changes by examining the diversity of Sec/

Cys use in humans. A few populations are known to currently inhabit regions of Se deficiency, whereas others are in regions of borderline Se toxicity (Levander 1987), but the question of whether Se geographical distribution has historically shaped human diversity is better served by an unbiased sample of human populations, which provide a cross-section of Se nutritional histories in the world (supplementary table S5, Supplementary Material online). All Sec and homologous Cys sites in the human genome were genotyped in the HGDP-CEPH panel (Cann et al. 2002) and no variation was found. Neutrality cannot be rejected as an explanation for the absence of variants ($P = 0.83$ by analytical method, $P \geq 0.99$ by coalescent simulations) due to power limitations given the population sample size, the small number of Sec and Cys sites, and the little average diversity of the human genome. Nevertheless, the absence of polymorphism observed suggests that natural variation in these sites is rare, if at all present, in human populations. This is consistent with a minor role for Se availability in shaping Sec use in human proteins.

Discussion

A fundamental question in Se biology is the extent of functional exchangeability between Sec and Cys amino acids, a measure of the distinct contribution of Sec to protein function. Sec is a nonstandard amino acid, and previous evolutionary studies on amino acid exchangeability have not considered this rare residue. To gain insight into this question, we have characterized the evolutionary forces shaping the exchange of Sec/Cys residues in vertebrates, a challenging inference given the small number of Sec sites in vertebrate proteomes. We believe this approach to be superior to physicochemical or experimental measures of exchangeability (Grantham 1974; Miyata et al. 1979; Yampolsky and Stoltzfus 2005) for the question at hand, as it discerns selection from mutational biases and it accounts for different fitness effects due to the use of a Se-dependent amino acid in proteins. The recent characterization of vertebrate selenoproteomes is believed to be quite complete (Kryukov et al. 2003; Castellano et al. 2004, 2005; Shchedrina et al. 2007), and the knowledge of the vertebrate phylogeny and mutation rates enables us, for the first time, to test current hypothesis on the role of Se and Sec in protein activity.

Our results are consistent not only with strong purifying selection acting on both Sec and Cys sites (as expected from functional sites), but also with a low level of functional exchangeability between the two residues over half a billion years of vertebrate evolution. These results underscore the unique role of Sec in protein activity. In interpreting these findings, it is worth noting that, as any evolutionary inference, they depend on the null model adopted and the test statistic used. In our simulations of the neutral divergence of vertebrate selenoproteomes, the expected number of synonymous substitutions per synonymous site is used as a proxy of the neutral mutation rate (see Materials and Methods). Synonymous mutations in mammals and other vertebrates with small population sizes are commonly assumed to be neutral. Although many syn-

onymous mutations are no doubt free from selection, selective pressures related to translational efficiency, mRNA stability, splicing control, and others suggest that weakly purifying selection may act on an unknown fraction of synonymous sites (Chamary et al. 2006). Weakly purifying selection would make us underestimate mutation rates in vertebrate genomes, but would not compromise the tests. On the contrary, a slower neutral rate of evolution would make our tests conservative in the inference of purifying selection, a statistical property shared by our divergence summary statistic (see Materials and Methods).

A more problematic bias would be the underestimation of the extent of mutation rate heterogeneity in a genome, which would result in an overestimation of sequence divergence. Such biased neutral expectation could result in the false inference of constraint. Several lines of evidence, though, suggest that this is an unlikely explanation to our results. First, synonymous sites in selenoproteins and Cys-homolog genes between humans and chimpanzees are not unusually constrained, suggesting that mutations accumulate at a typical rate in these genes (Castellano S, data not shown). Second, selenoproteins and Cys-homolog genes are located in different chromosomal regions within and between species genomes. That the distribution of a large fraction of these genes consistently overlaps regions of low mutation in most species, as the pervasive purifying selection inferred above would imply, is highly improbable. Third, neutral simulations with increasing levels of mutation rate heterogeneity suggest that our tests are, to a large degree, robust to nonuniform mutation rates. Therefore, all evidence supports that vertebrate selenoproteomes are selectively constrained and that such evolutionary conservation can be of functional relevance.

Accordingly, we discuss previously proposed selective pressures on Sec usage in the context of the inferred constraint:

1. Nutrition is a prominent selective force in humans and other species (Haygood et al. 2007), and dietary adaptations are likely to have arisen primarily due to changes in nutrient availability. For example, iron deficiency in populations of European descent may have caused recent local positive selection on the *HFE* gene (iron absorption regulation), where an enhancing Cys to Tyr mutation has reached a relatively high frequency in only ~60 generations (Bamshad and Wooding 2003). Environmental changes and range expansions in populations may also have resulted in different nutritional pressures regarding Se dietary intake, an unevenly distributed trace element worldwide (Shamberger 1981; Levander 1987; Valentine 1997). Indeed, selective claims regarding Se availability in vertebrates and other eukaryotes have been recently published (Lobanov et al. 2007, 2008a, 2008b). If so, patterns of selenoproteome divergence and diversity should bear the footprint of past and present Se abundance or deficiency events. Despite the fact that vertebrate species may have repeatedly encountered extreme Se environments in the last half billion years, our exchangeability test fails to support extensive positive selection targeting Sec/Cys sites (fig. 2A). This

result is consistent with low functional exchangeability between Sec and Cys amino acids and a minor role for environmental Se in driving the use of Sec in vertebrate enzymes. Furthermore, despite a considerable range of variation in dietary Se intake among human populations (Levander 1987), we find no evidence of variation in the use of Sec and Cys residues among populations worldwide (see supplementary table S5, Supplementary Material online), suggesting that Se availability has not sized the human selenoproteome among regions throughout the world.

2. Atmospheric O_2 levels have played a key role in the evolution of vertebrates (Canfield et al. 2007). Leinfelder et al. (1988) have suggested that the highly oxidizable Sec (Jacob et al. 2003) is counterselected (substituted by Cys) in response to rising O_2 levels, a hypothesis later embraced by Jukes (1990). Although this adaptive factor was suggested to be important 2.4 billion years ago, examples of molecular adaptations to variable O_2 concentrations have been described in animals (Bargelloni et al. 1998). A great increase in O_2 levels in late Proterozoic (~600 Ma) preceded the appearance of the first animals, and wide variations in atmospheric O_2 concentrations followed in the Phanerozoic (~550 Ma to the present). Vertebrates have evolved for half a billion years with a maximum O_2 concentration around 300 Ma (~31% O_2), a minimum about 200 Ma (~13% O_2), followed by a steady rise to present times (21% O_2) (Berner 2006; Berner et al. 2007). O_2 levels have been recently proposed to drive nonneutral evolution of eukaryotic selenoproteomes (Lobanov et al. 2007, 2008a, 2008b). The extensive constraint identified in Sec and Cys residues during vertebrate evolution (fig. 2A and B) is, however, in agreement with a limited role of O_2 in shaping Sec usage, as broad fluctuations in selection intensity would have resulted in episodic positive selection, most likely in different genes in different lineages, leading to higher selenoproteome divergence. In agreement, no significant negative correlation between O_2 levels and selenoproteomes sizes during the phanerozoic was found (Results and supplementary table S4, Supplementary Material online). However, the uncertainty of these estimates, particularly of divergence times between lineages, and the small number of selenoproteomes tested, makes this lack of correlation, at most, suggestive. Nevertheless, the observation that vertebrate selenoproteomes have remained similar in size, virtually unchanged in mammals, for hundreds of millions of years despite levels of atmospheric O_2 exhibiting the greatest variability of any geological period, is a stronger evidence of a minor role of O_2 concentrations in driving Sec use in vertebrates.
3. Metabolic costs of amino acid biosynthesis and incorporation into proteins are usually overlooked selective pressures (Akashi and Gojobori 2002). Sec is an expensive residue due to its complex biosynthetic pathway (Xu et al. 2007) and its elaborate and inefficient cotranslational insertion into proteins (Berry et al. 1992; Driscoll and Copeland 2003; Mehta et al. 2004). However, the strong purifying selection in both Cys and, more importantly, Sec sites (fig. 2A and B) suggests no major detrimental effect on fitness of Sec larger metabolic cost. Other than Sec anabolic fitness effects, the slightly higher number of Sec to Cys than Cys to Sec changes can be attributed to the requirement of a functional SECIS element in selenoproteins. This result provides some support, at least in vertebrates, to the pattern of Sec usage following Dollo's Principle (Farris 1977), in which the derived state (Sec) arose only once and reversals to Cys have occurred multiple times.
4. Functional constraints on particular amino acid sites, although difficult to document, can explain in part heterogeneity in protein rates of evolution. The extent of constraint in Sec and Cys sites across vertebrate selenoproteomes strongly suggests that some functional characteristics account for the low exchangeability between Sec and Cys residues (fig. 2A and B). The fine molecular features behind the observed degree of constraint in each selenoprotein or Cys homolog may vary and are not fully clear, as the majority of these enzymes remain poorly characterized. Nevertheless, it is now apparent that the higher catalytic activity usually attributed to Sec-containing enzymes (Berry et al. 1992; Rocher et al. 1992; Maiorino et al. 1995; Zhong and Holmgren 2000) can only justify a fraction of the extensive conservation in Sec and Cys sites during vertebrate evolution. Similar catalytic activity between homologous Sec- and Cys-containing enzymes, most likely due to additional compensatory substitutions in the active site of Cys enzymes, has been recently reported (Gromer et al. 2003; Kim and Gladyshev 2005; Shchedrina et al. 2007). A broader range of substrates and pH in which selenoenzyme activity is possible (Gromer et al. 2003) or different catalytic mechanisms between Sec- and Cys enzymes (Kim and Gladyshev 2005) may account for the constraint and the deleterious effect of Sec/Cys replacements inferred here. A more complex view of Sec in protein activity is emerging, and other biochemical and functional differences with fitness consequences may apply to the majority of uncharacterized selenoenzymes. Hence, to the question posed by Johansson et al. (2005) of whether every reaction catalyzed by Sec can be supported by Cys, the evolutionary analysis of all Sec and Cys residues in vertebrate proteomes provides a negative answer. Overall, our results support and extend to the protein, organismal, and population level the characterized physicochemical differences between Se and S (Stadtman 1996).

We have derived a global measure of functional exchangeability across vertebrate selenoproteins and selenoproteomes and provided the first evolutionary assessment of several selective pressures proposed to drive Sec use in proteins. The low exchangeability between Sec and Cys residues is better explained by strong natural selection due to Sec/Cys functional differences and, at best, a moderate role of environmental and metabolic forces, suggesting caution in the interpretation of evolutionary trends in Sec usage as ecological adaptations. Although our results only apply to

the vertebrate clade, we feel that common claims of ecological adaptations in the Se field may be premature. Despite the difficulties and uncertainties associated with any molecular inference of the past, different selective factors leave different signatures of selection and these adaptive hypotheses can be examined through established evolutionary principles. Strong evidence for selection is most needed for genes of plausible ecological importance, like selenoproteins, as apparent selective factors may discourage considering alternatives to environmental adaptations (Gould and Lewontin 1979; Mitchell-Olds et al. 2007). Furthermore, natural selection is just one of several evolutionary mechanisms responsible for differences at the molecular level (Lynch 2007) and, despite typical assumptions in Se biology regarding the role of natural selection, no Sec to Cys or Cys to Sec substitution has yet been shown to be adaptive. Whether nonneutral evolutionary processes are responsible for some of these amino acid replacements is unknown. Similarly, whether adaptation to local Se levels or other selective factors have driven the evolution of selenoprotein expression, Se intake, metabolism or transport has not been addressed. These are open questions in Se biology.

A better understanding of the selenoproteomes and neutral evolutionary patterns in other taxa will be necessary to fully assess the generality of our conclusions. For example, the recent identification in the *Drosophila* clade of the first animal without selenoproteins is remarkable (*Drosophila* 12 Genomes Consortium 2007). Although all known *Drosophila* species have three selenoproteins, *Drosophila willistoni* has none. Indeed, insects seem to have a higher number of Sec/Cys exchanges in proteins than vertebrates (Chapple and Guigó 2008; Lobanov et al. 2008a, 2008b). The evolutionary forces and selective pressures, if any, driving these replacements are still unclear. Beyond the Sec residue, the evolutionary forces targeting selenoprotein genes as a whole are also poorly known. A notable exception is the Glutathione peroxidase 1 gene, which may have been under adaptive evolution in recent human history (Foster et al. 2006). In any case, if the results obtained here are representative of more divergent species, the certain conclusion is the unique role of Sec in protein activity and evolution. Overall, Sec and Cys residues may be less functionally exchangeable than usually thought and, if some instances of Sec/Cys substitutions have been adaptive in vertebrates or other taxa, Sec distinct biochemical properties, and not Se geographical distribution, global O₂ levels nor metabolic cost, may have played a major role in the evolution of selenoproteomes.

Supplementary Material

Supplementary figures S1 and S2 and tables S1–S7 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

S.C. thanks S.R. Eddy for time and resources to complete this manuscript. We thank M.J. Berry for helpful comments and suggestions; E.H. Margulies for sharing

unpublished data on vertebrate rates of neutral evolution; R.A. Berner for providing up-to-date estimates of atmospheric O₂ in the Phanerozoic eon; and M. Vallés for technical assistance. This work was supported by grants BIO2006-03380 from the Spanish Ministry of Education and Biosapiens LSHG-CT-2003-503265 from the European Commission (FP6 Program) (to R.G.) and NIH GM065509 (to A.G.C.).

Literature Cited

- Akashi H, Gojobori T. 2002. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc Natl Acad Sci USA*. 99:3695–3700.
- Bamshad M, Wooding SP. 2003. Signatures of natural selection in the human genome. *Nat Rev Genet*. 4:99–111.
- Bargelloni L, Marcato S, Patarnello T. 1998. Antarctic fish hemoglobins: evidence for adaptive evolution at subzero temperature. *Proc Natl Acad Sci USA*. 95:8670–8675.
- Berner RA. 2006. GEOCARBSULF: a combined model for Phanerozoic atmospheric O₂ and CO₂. *Geochim Cosmochim Acta*. 70:5653.
- Berner RA, VandenBrooks JM, Ward PD. 2007. Oxygen and evolution. *Science*. 316:557–558.
- Berry MJ, Mai AL, Kieffer J, Harney JW, Larsen P. 1992. Substitution of cysteine for selenocysteine in type I iodothyronine deiodinase reduces the catalytic efficiency of the protein but enhances its translation. *Endocrinology*. 131:1448–1452.
- Canfield DE, Poulton SW, Narbonne GM. 2007. Late-Neoproterozoic deep-ocean oxygenation and the rise of animal life. *Science*. 315:92–95.
- Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A. 2002. A human genome diversity cell line panel. *Science*. 296:261–262.
- Castellano S. 2009. On the unique function of selenocysteine – insights from the evolution of selenoproteins. *Biochim Biophys Acta*. Advanced online publication. doi:10.1016/j.bbagen.2009.03.027.
- Castellano S, Gladyshev VN, Guigó R, Berry MJ. 2008. SelenoDB 1.0: a database of selenoprotein genes, proteins and SECIS elements. *Nucl Acids Res*. 36:D339–D343.
- Castellano S, Lobanov AV, Chapple C, et al. (11 co-authors). 2005. Diversity and functional plasticity of eukaryotic selenoproteins: identification and characterization of the SelJ family. *Proc Natl Acad Sci USA*. 102:16188–16193.
- Castellano S, Novoselov SV, Kryukov GV, Lescure A, Blanco E, Krol A, Gladyshev VN, Guigó R. 2004. Reconsidering the evolution of eukaryotic selenoproteins: a novel nonmammalian family with scattered phylogenetic distribution. *EMBO Rep*. 5:71–77.
- Chamary JV, Guigó R, Parmley JL, Hurst LD. 2006. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat Rev Genet*. 7:98–108.
- Chapple CE, Guigó R. 2008. Relaxation of selective constraints causes independent selenoprotein extinction in insect genomes. *PLoS ONE*. 3:e2968.
- Driscoll DM, Copeland DR. 2003. Mechanism and regulation of selenoprotein synthesis. *Annu Rev Nutr*. 23:17–40.
- Drosophila* 12 Genomes Consortium. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*. 450:203–218.
- Ellegren H, Smith NG, Webster MT. 2003. Mutation rate variation in the mammalian genome. *Curr Opin Genet Dev*. 13:562–568.

- Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet.* 8:610–618.
- Farris JS. 1977. Phylogenetic analysis under Dollo's law. *Syst Zool.* 26:77–88.
- Fitch WM. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. *Syst Zool.* 20:406–416.
- Foster CB, Aswath A, Chanock SJ, McKay HF, Peters U. 2006. Polymorphism analysis of six selenoprotein genes: support for a selective sweep at the glutathione peroxidase 1 locus (3p21) in Asian populations. *BMC Genet.* 7:56.
- Gish W, States DJ. 1993. Identification of protein coding regions by database similarity search. *Nat Genet.* 3:266–272.
- Gould SJ, Lewontin RC. 1979. The spandrels of San Marco and the Panglossian paradigm. *Proc R Soc Lond.* 205:581–598.
- Grantham R. 1974. Amino acid difference formula to help explain protein evolution. *Science.* 185:862–864.
- Gromer S, Johansson L, Bauer H, Arscott LD, Rauch S, Ballou DP, Williams CH Jr, Schirmer RH, Arnér ES. 2003. Active sites of thioredoxin reductases: why selenoproteins? *Proc Natl Acad Sci USA.* 100:12618–12623.
- Hasegawa M, Kishino H, Yano T. 1985. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol.* 22:160–174.
- Haygood R, Fedrigo O, Hanson B, Yokoyama KD, Wray GA. 2007. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. *Nat Genet.* 39:1140–1144.
- Hedges SB. 2002. The origin and evolution of model organisms. *Nat Rev Genet.* 3:838–849.
- Hein J, Schierup MH, Wiuf C. 2005. From genealogies to sequences. In: *Gene genealogies, variation and evolution. A primer in coalescent theory.* Oxford: Oxford University Press. pp 57.
- Hudson RR. 2002. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics.* 18:337–338.
- Jacob G, Giles GI, Giles NM, Sies NH. 2003. Sulfur and selenium: the role of oxidation state in protein structure and function. *Angew Chem Int Ed.* 42:4742–4758.
- Johansson L, Gafvelin G, Arnér ES. 2005. Selenocysteine in proteins—properties and biotechnological use. *Biochim Biophys Acta.* 1726:1–13.
- Jukes TH. 1990. Genetic code 1990. *Experientia.* 46:1149–1157.
- Kanzok SM, Fechner A, Bauer H, Ulschmid JK, Müller HM, Botella-Muñoz J, Schneuwly S, Schirmer R, Becker K. 2001. Substitution of the thioredoxin system for glutathione reductase in *Drosophila melanogaster*. *Science.* 291:643–646.
- Kawashima S, Pokarowski P, Pokarowska M, Kolinski A, Katayama T, Kanehisa M. 2008. AAindex: amino acid index database, progress report 2008. *Nucleic Acids Res.* 36:D202–D205.
- Kim HY, Gladyshev VN. 2005. Different catalytic mechanisms in mammalian selenocysteine- and cysteine-containing methionine-R-sulfoxide reductases. *PLoS Biol.* 3:e375.
- Kryukov GV, Castellano S, Novoselov SV, Lobanov AV, Zettab O, Guigó R, Gladyshev VN. 2003. Characterization of mammalian selenoproteomes. *Science.* 300:1439–1443.
- Leinfelder W, Zehlein E, Mandrand-Berthelot M, Böck A. 1988. Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. *Nature.* 331:723–725.
- Levander OA. 1987. A global view of human selenium nutrition. *Ann Rev Nutr.* 7:227–250.
- Lobanov AV, Fomenko DE, Zhang Y, Sengupta A, Hatfield DL, Gladyshev VN. 2007. Evolutionary dynamics of eukaryotic selenoproteomes: large selenoproteomes may associate with aquatic life and small with terrestrial life. *Genome Biol.* 8:R198.
- Lobanov AV, Hatfield DL, Gladyshev VN. 2008a. Reduced reliance on the trace element selenium during evolution of mammals. *Genome Biol.* 9:R62.
- Lobanov AV, Hatfield DL, Gladyshev VN. 2008b. Selenoproteinless animals: selenophosphate synthetase SPS1 functions in a pathway unrelated to selenocysteine biosynthesis. *Protein Sci.* 17:176–182.
- Lynch M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc Natl Acad Sci USA.* 104:8597–8604.
- Maiorino M, Aumann KD, Brigelius-Flohé R, Doria D, van den Heuvel J, McCarthy J, Roveri A, Ursini F, Flohé L. 1995. Probing the presumed catalytic triad of selenium-containing peroxidases by mutational analysis of phospholipid hydroperoxide glutathione peroxidase (PHGPx). *Biol Chem Hoppe Seyler.* 376:651–660.
- Margulies EH, Vinson JP, NISC Comparative Sequencing Program, et al. (11 co-authors). 2005. An initial strategy for the systematic identification of functional elements in the human genome by low-redundancy comparative sequencing. *Proc Natl Acad Sci USA.* 102:3354–3359.
- Marth GT, Czebark E, Murvai J, Sherry ST. 2004. The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics.* 166:351–372.
- Mehta A, Rebsch CM, Kinzy SA, Fletcher JE, Copeland PR. 2004. Efficiency of mammalian selenocysteine incorporation. *J Biol Chem.* 279:37852–37859.
- Mitchell-Olds T, Willis JH, Goldstein DB. 2007. Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nat Rev Genet.* 8:845–856.
- Miyata T, Miyazawa S, Yasunaga T. 1979. Two types of amino acid substitutions in protein evolution. *J Mol Evol.* 12:219–236.
- Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG. 2007. Recent and ongoing selection in the human genome. *Nat Rev Genet.* 8:857–868.
- Nikolaev S, Montoya-Burgos JI, Margulies EH, ISC Comparative Sequencing Program, Rougemont J, Nyffeler B, Antonarakis SE. 2007. Early history of mammals is elucidated with the ENCODE multiple species sequencing data. *PLoS Genet.* 3:e2.
- Rambaut A, Grassly NC. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput Appl Biosci.* 13:235–238.
- Rocher C, Lalanne JL, Chaudière J. 1992. Purification and properties of a recombinant sulfur analog of murine selenium–glutathione peroxidase. *Eur J Biochem.* 205:955–960.
- Rosenberg MS, Subramanian S, Kumar S. 2003. Patterns of transitional mutation biases, within and among mammalian genomes. *Mol Biol Evol.* 20:988–993.
- Shamberger RJ. 1981. Selenium in the environment. *Sci Total Environ.* 17:59–74.
- Shchedrina VA, Novoselov SV, Malinouski MY, Gladyshev VN. 2007. Identification and characterization of a selenoprotein family containing a diselenide bond in a redox motif. *Proc Natl Acad Sci USA.* 104:13919–13924.
- Stadtman TC. 1996. Selenocysteine. *Ann Rev Biochem.* 65:83–100.
- Valentine JL. 1997. Environmental occurrence of selenium in waters and related health significance. *Biomed Environ Sci.* 10:292–299.
- Venter JC, Adams MD, Myers EW, et al. (275 co-authors). 2001. The sequence of the human genome. *Science.* 291:1304–1351.

- Williamson SH, Hernandez R, Fledel-Alon A, Zhu L, Nielsen R, Bustamante CD. 2005. Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proc Natl Acad Sci USA*. 102:7882–7887.
- Xu XM, Carlson BA, Mix H, Zhang Y, Saira K, Glass RS, Berry MJ, Gladyshev VN, Hatfield DL. 2007. Biosynthesis of Selenocysteine on Its tRNA in Eukaryotes. *PLoS Biol*. 5:e4.
- Yampolsky LY, Stoltzfus A. 2005. The exchangeability of amino acids in proteins. *Genetics*. 170:1459–1472.
- Zhong L, Holmgren A. 2000. Essential role of selenium in the catalytic activities of mammalian thioredoxin reductase revealed by characterization of recombinant enzymes with selenocysteine mutations. *J Biol Chem*. 275:18121–18128.
- Zmasek CM, Eddy SR. 2001. A simple algorithm to infer gene duplication and speciation events on a gene tree. *Bioinformatics*. 17:821–828.

Manolo Gouy, Associate Editor

Accepted May 22, 2009